

# Optimizing Gaze Direction in a Visual Navigation Task

Tuomas Välimäki and Risto Ritala

**Abstract**—Navigation in an unknown environment consists of multiple separable subtasks, such as collecting information about the surroundings and navigating to the current goal. In the case of pure visual navigation, all these subtasks need to utilize the same vision system, and therefore a way to optimally control the direction of focus is needed. We present a case study, where we model the active sensing problem of directing the gaze of a mobile robot with three machine vision cameras as a partially observable Markov decision process (POMDP) using a mutual information (MI) based reward function. The key aspect of the solution is that the cameras are dynamically used either in monocular or stereo configuration. The benefits of using the proposed active sensing implementation are demonstrated with simulations and experiments on a real robot.

## I. INTRODUCTION

In human vision, the sequential deployment of gaze in multiple directions is a vital part of information gathering [1]. If a person is given a task to follow a set of driving instructions, they seemingly effortlessly navigate to the desired destination, while simultaneously observing their surroundings for possible landmarks or obstacles. Similar active sensing approaches are used in mobile robotics for e.g. sensor selection. However, there is little research done on actively manipulating individual sensors in a vision system to optimize information gathering.

Active sensing in general refers to seeking a policy for determining the optimal sensor configuration at each time instance as a function of information from previous measurements to achieve a goal [2]. The problem can crudely be categorized as an information-gathering problem or as a task-achievement problem depending on whether the goal is to gather maximum amount of information or merely the completion of a task not related to sensing itself. We focus only on the former.

Robots face uncertainty both in predicting and sensing their own state and that of the surroundings. Regardless of whether the goal is to gather information or achieve a task, the active sensing problem becomes a sequential decision making problem in a stochastic environment and can therefore be modeled as a partially observable Markov decision process (POMDP) [3]. The utility of actions is measured by a reward function and the robot acts so as to maximize the expected reward. Because the state is not directly observable, the robot's knowledge of it is described as a probability density function (pdf) over the state known as the belief state.

T. Välimäki and R. Ritala are with the Department of Automation Science and Engineering, Tampere University of Technology, P.O. Box 692, FI-33101, Tampere, Finland. Email: [tuomas.s.valimaki@tut.fi](mailto:tuomas.s.valimaki@tut.fi), [risto.ritala@tut.fi](mailto:risto.ritala@tut.fi)

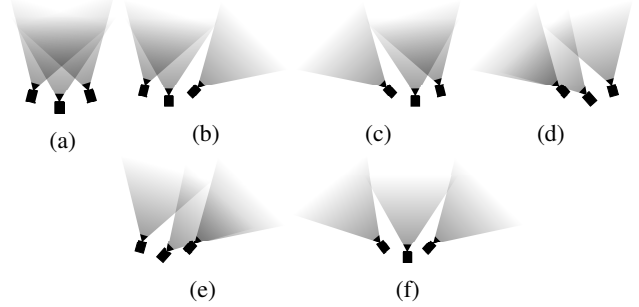


Fig. 1: The 6 camera modes of the proposed system.

Typical information-gathering problems in robotics are exploration tasks. Controlling the direction of sensor focus independently of the robot pose is beneficial when navigating in unknown environments, especially on car-like robots that cannot turn on spot, to gather information about the surroundings. Nevertheless, often only the pose of the robot is considered, and the sensor placements are fixed. In [4] a solution for integrated simultaneous localization and mapping (SLAM) and exploration is proposed, using a Rao-Blackwellized particle filter to compute the expected information gain of actions. The approach can be regarded as an approximation of a myopic POMDP.

Often the active sensing problems are modeled as task-achievement problems, where the reward function depends only on state and actions and is therefore linear in the belief state and can be solved using standard POMDP solvers. Such solutions have been presented e.g. in [5], [6]. In pure information-gathering problems it is sometimes convenient to use information theoretic rewards that are nonlinear in the belief state. For example [7] mentions several rewards typically used for maximizing information, such as the Kullback-Leibler (KL) divergence, which has also been shown to explain human visual attention [8]. Other nonstandard POMDPs have also been proposed e.g. [9], [10].

Recently [11] discusses an active sensing problem, where a robot must decide the direction of focus of its vision system while moving along a fixed trajectory, and proposes a solution via an approximate open-loop feedback control. We study a similar case where the trajectory is not fixed, but consists of landmarks at uncertain locations motivated by driving instructions for human operators; drive straight until you reach a yellow house, turn left and continue a few kilometers until reaching a blue villa.

Our robot must select the direction of gaze of its vision system while navigating a sequence of visual landmarks in order to maximize the amount of information about the

positions of the current and next landmark in sequence. As opposed to curiosity driven exploration (e.g. [10]), the landmarks represent navigational goals for an underlying path planner. We formulate the problem as a myopic POMDP with information-theoretic reward and solve it online.

The vision system consists of three cameras. The cameras can be in 6 discrete modes as seen in Fig. 1 to either maximize the field of view or to focus two or even three cameras in the same direction. When the fields of view of any two cameras overlap, they are used as a stereo pair to allow depth computation as opposed to a monocular camera, where it is only possible to measure the relative angle of the landmark seen in the image w.r.t. the camera. Using the cameras in a monocular configuration is however beneficial when searching for a landmark to narrow down its possible location, which can then be refined by using a stereo pair. The main contribution of this paper is to demonstrate the benefits of using the cameras in a dynamic monocular or stereo configurations depending on the expected information gain, as opposed to a fixed stereo configuration.

The remainder of the paper is organized as follows. In Section II we outline the basic principles of partially observable Markov decision processes. Section III formulates the case problem and presents the solution. The results of the experiments are reported in Section IV, and Section V concludes the paper.

## II. PRELIMINARIES

Partially observable Markov decision processes (POMDPs) are formally presented as the tuple  $\langle S, A, T, R, Z, O \rangle$ . Consider a sequential decision process, where  $s \in S$  denotes the state of the system at any time  $t$ . At each time step the agent takes an action  $a \in A$  and receives a reward  $R : S \times A \rightarrow \mathbb{R}$ , where  $R(s, a)$  is the expected reward of executing action  $a$  in state  $s$ .  $T : S \times A \times S \rightarrow [0, 1]$  is the state transition model, where  $T(s, a, s') = p(s'|s, a)$  describes the probability of ending in state  $s'$  when taking the action  $a$  in state  $s$ . In this model, the next state and the expected reward depend only on the current state and the action taken, which is called the *Markov* property. This type of decision processes are called Markov decision processes (MDPs) [12]–[14].

In case the observations about the state are imperfect and noisy, we model the problem as a POMDP [15], [16]. Let us define  $Z$  as the set of all possible observations and  $O : S \times A \times Z \rightarrow [0, 1]$  as the observation function, where  $O(s', a, z') = p(z'|a, s')$  is the probability of observing  $z'$  if action  $a$  is performed and the resulting state is  $s'$ .

Because the states are not directly observable, the agent cannot choose its actions based on the state only. Instead it has to consider the whole history of actions and states  $h_t = \{a_0, z_1, \dots, a_{t-1}, z_t\}$ . Maintaining all this information would be memory expensive, but the belief state  $b(s) = p(s_t = s | h_t, b_0)$ , where  $b_0$  is the initial belief about the starting state, is a sufficient statistic for the history [15].

The next belief state  $b'$  can be calculated from the previous belief state  $b$ , previous action  $a$ , and current observation  $z'$

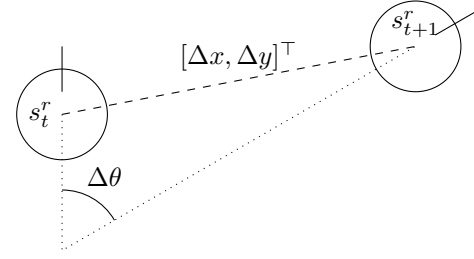


Fig. 2: Illustration of robot movement between time instances  $t$  and  $t + 1$ .

using the belief update function  $\tau(b, a, z')$  defined as

$$b'(s') = \tau(b, a, z')(s') = \frac{1}{p(z'|b, a)} O(s', a, z') \sum_{s \in S} T(s, a, s') b(s), \quad (1)$$

where  $p(z'|b, a)$  is the prior probability of observing  $z'$  and acts as the normalizing term.

The objective of POMDP planning is to find actions that maximize the obtained reward over a time horizon and results in a policy that maps all belief states  $b \in B$  to actions;  $\pi : B \rightarrow A$ . Instead of defining a reward that depends on the state, we define an information theoretic reward  $R_b(b, a)$  that depends directly on the belief state. See e.g. [7]. As we consider only solutions over one time step i.e. act greedily w.r.t. the expected immediate reward, the optimal policy is

$$\pi^* = \arg \max_{\pi} \mathbb{E} [R_b(b, \pi(b))], \quad (2)$$

where  $R_b(b, \pi(b))$  is the immediate reward for executing policy  $\pi$  in belief state  $b$ .

## III. PROBLEM FORMULATION

### A. Case description

The robot has to select the direction of gaze of its vision system while following a given sequence of landmarks. The *a priori* information of each landmark position is a highly uncertain multivariate Gaussian distribution. The task can be divided into two separate subtasks: 1) maximizing information about the position of the current landmark in sequence, and 2) maximizing information about the position of the next landmark in sequence. The robot moves at a constant velocity and both subtasks are solved to maximize the probability that the robot actually visits all of the landmarks. When the robot believes it is close enough to the current landmark, it starts navigating to the next.

The three cameras on the robot can be utilized, as shown in Fig. 1, so that either (a) all cameras face forward, (b) left and middle cameras face forward while the right camera faces right, (c) right and middle cameras face forward while the left camera faces left, (d) left and middle cameras face left while the right camera faces forward, (e) right and middle cameras face right while the left camera faces forward, or (f) all cameras face different directions. When the fields of view of any two cameras overlap, they are used as a stereo pair.

### B. Dynamics and observation models

The robot's state  $s_t^r$  at time instant  $t$  is given by its location  $x_t$ ,  $y_t$ , and its heading  $\theta_t$ . The robot is controlled by its translational velocity  $v_t$ , and its rotational velocity  $\omega_t$ . The robots relative movement between time instances  $t$  and  $t+1$ , as depicted in Fig. 2, is calculated from the control signals as follows

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta \theta \end{bmatrix} = \begin{bmatrix} v_t \Delta t \cos(\omega_t \Delta t) \\ -v_t \Delta t \sin(\omega_t \Delta t) \\ \omega_t \Delta t \end{bmatrix}. \quad (3)$$

The control is affected by additive Gaussian noise, so that the joint pdf of  $u_t = [v_t, \omega_t]^\top$  is  $\mathcal{N}(\hat{u}_t, Q)$ , where  $\hat{u}_t$  are the desired control inputs and  $Q = \text{diag}(\sigma_v^2, \sigma_\omega^2)$  is the noise covariance.

The robot is set to move forward with a constant translational velocity,  $v_t = v$ , until it believes to be close enough to the last landmark in sequence. The rotational velocity for each time step,  $\omega_t$ , is calculated with a discrete time PID controller that minimizes the perpendicular distance of the robot from the desired path—a sequence of linear segments between the landmarks.

The  $N$  stationary landmarks are positioned at locations  $l^i = \{l_i^x, l_i^y\}_{i=1}^N$  and the system state is defined as the coordinates of each landmark in the robot's coordinate frame,  $s_t = [l_t^1, l_t^2, \dots, l_t^N]^\top$ . The transition model  $l_{t+1}^i = f(l_t^i, u_t)$  for each landmark can be written as

$$l_{t+1}^i = \begin{bmatrix} \cos(\Delta \theta) & \sin(\Delta \theta) \\ -\sin(\Delta \theta) & \cos(\Delta \theta) \end{bmatrix} l_t^i - \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}. \quad (4)$$

The robot can observe landmarks that reside inside any of the cameras' cone of observation  $C_j(a_t, \alpha, r_{max})_{j=1:3}$  defined by the action  $a_t \in A$ , view angle  $\alpha$ , and maximum range  $r_{max}$ . The action space  $A$  is the six camera modes presented in Fig. 1. The camera mode determines which cameras are operated in stereo and which in monocular mode.

In stereo mode the observation consists of the measured landmark coordinates  $l_t^i$  and the measurement model  $z_t = h_{stereo}(l_t^i) + w_t^{stereo}$  can be written

$$z_t = l_t^i + w_t^{stereo}, \quad (5)$$

where  $w_t^{stereo}$  is Gaussian noise with zero mean and covariance  $W^{stereo} = \text{diag}(\sigma_x^2, \sigma_y^2)$ . In monocular mode we can only measure the angle  $\beta$  of the landmark relative to the robot's coordinate frame and the measurement model  $z_t = h_{mono}(l_t^i) + w_t^{mono}$  is

$$z_t = \beta_t + w_t^{mono} = \arctan \left( \frac{l_t^y}{l_t^x} \right) + w_t^{mono}, \quad (6)$$

where  $w_t^{mono}$  is Gaussian noise with zero mean and variance  $W^{mono} = \sigma_\beta^2$ .

### C. State estimation

We apply an extended Kalman filter (EKF) to track the robot's belief about the landmark locations,  $L_t^i|_{i=1:N}$ , which are assumed to be independent random variables. Given the

previous belief  $L_{t-1}^i \sim b_{t-1}^i = \mathcal{N}(\mu_{t-1}^i, \Sigma_{t-1}^i)$ , at each timestep we first calculate the prediction on the following time step  $L_t^{i+} \sim b_t^{i+} = \mathcal{N}(\mu_t^{i+}, \Sigma_t^{i+})$  according to

$$\mu_t^{i+} = f(\mu_{t-1}^i, u_{t-1}) \quad (7a)$$

$$\Sigma_t^{i+} = F_s \Sigma_{t-1}^i F_s^\top + F_u Q F_u^\top, \quad (7b)$$

where  $F_s$  and  $F_u$  are Jacobians of the transition function  $f$  w.r.t. state and control respectively, evaluated at  $(\mu_{t-1}^i, u_{t-1})$ .

Every time a measurement is received, the posterior  $L_t^i|a_{t-1} \sim b_t^i = \mathcal{N}(\mu_t^i, \Sigma_t^i)$  is calculated according to

$$d_t = z_t - h(\mu_t^{i+}) \quad (8a)$$

$$E_t = H_s \Sigma_t^{i+} H_s^\top + W \quad (8b)$$

$$K_t = \Sigma_t^{i+} H_s^\top E_t^{-1} \quad (8c)$$

$$\mu_t^i = \mu_t^{i+} + K_t d_t \quad (8d)$$

$$\Sigma_t^i = \Sigma_t^{i+} - K_t E_t K_t^\top, \quad (8e)$$

where  $H_s$  is the Jacobian of the measurement function  $h$  w.r.t. state. Depending which measurement is received, the appropriate measurement function  $h_{stereo}$  or  $h_{mono}$  is used.

### D. Reward function

As the task consist of gaining information about the current and next landmarks in sequence, we need a way to measure the expected information gain. Information theoretic quantities such as the Kullback-Leibler (KL) divergence [17] are often used to measure the information gain between two pdfs. The expected KL divergence can be expressed as the mutual information (MI) of the random variables.

We want to maximize the MI,  $\mathcal{I}(L_{t+1}^{i+}, L_{t+1}^i|a_t)$ , of the predicted and expected landmark locations, so the total immediate reward is

$$R(b_t, a_t) = \gamma \mathcal{I}(L_{t+1}^{i+}, L_{t+1}^i|a_t) + \xi \mathcal{I}(L_{t+1}^{i+1}, L_{t+1}^{i+1}|a_t) - g(a_t), \quad (9)$$

where  $\gamma$  and  $\xi$  are weighing factors to prioritize gathering information either about the current or about the next landmark and  $g(a_t)$  is a cost for changing camera modes defined as

$$g(a_t) = \begin{cases} \zeta, & a_t \neq a_{t-1} \\ 0, & a_t = a_{t-1}, t = 0 \end{cases}, \quad (10)$$

where  $\zeta$  is a positive scalar constant.

MI of the predictive and expected landmark locations can be calculated as

$$\begin{aligned} \mathcal{I}(L_{t+1}^{i+}, L_{t+1}^i|a_t) &= \mathbb{E}_Z [D_{KL}(b_{t+1}^{i+}, b_{t+1}^i)] \\ &= \int_Z D_{KL}(b_{t+1}^{i+}, b_{t+1}^i) p(z_{t+1}|a_t, b_t^i) dz_{t+1} \end{aligned} \quad (11)$$

where  $b_{t+1}^i$  depends on the observation  $z_{t+1}$ , and  $p(z_{t+1}|a_t, b_t^i)$  is the probability that the landmark  $l_{t+1}^i$  is visible in observation  $z_{t+1}$  if action  $a_t$  is taken in current belief  $b_t^i$ .  $D_{KL}(\cdot)$  is the KL divergence (see e.g. [18]).

The probability of  $l_{t+1}^i$  being visible in observation  $z_{t+1}$  would be difficult to calculate, so instead we use Monte Carlo

methods and draw  $K$  samples from the probability distribution;  $\{z_{t+1}^k\}_{k=1}^K \sim p(z_{t+1}|a_t, b_t^i) = \mathcal{N}(h(\mu_{t+1}^{i+}), E_{t+1})$ . In other words, we sample possible landmark locations from  $\mathcal{N}(\mu_{t+1}^{i+}, \Sigma_{t+1}^{i+})$  and expect to get a measurement  $z_{t+1}^k$  from each sample that lies inside the appropriate cones of observation  $C_j(a_t, \alpha, r_{max})_{j=1:3}$ . The camera mode  $a_t$  determines which cameras are used as a stereo pair, i.e. the sample must be inside both cones of observation, and which as monocular cameras. Now (11) can be approximated with

$$\mathcal{I}(L_{t+1}^{i+}, L_{t+1}^i | a_t) \approx \frac{1}{K} \sum_{k=1}^K D_{KL}(b_{t+1}^{i+}, b_{t+1}^i), \quad (12)$$

where  $b_{t+1}^i$  depends on the sampled measurement  $z_{t+1}^k$ .

#### E. Solving the problem

Acting greedily may result in poorer performance if 1) the focus of attention cannot be changed rapidly, or 2) the observed features are not stationary [11]. As these do not apply to the problem defined above, we consider only myopic optimization of the immediate information gain.

At timestep  $t$  we solve

$$a_t^* = \arg \max_{a_t \in A} R(b_t, a_t) \quad (13)$$

to obtain the optimal action.

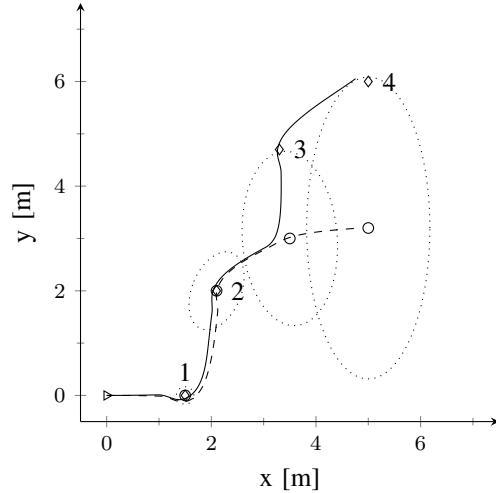
### IV. EXPERIMENTS

#### A. Simulation experiments

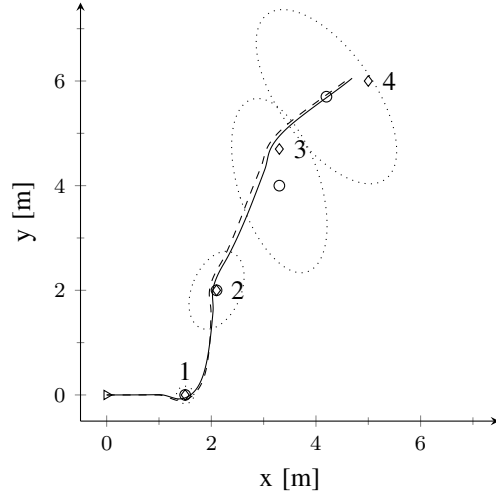
We demonstrate with simulation experiments the benefits of using the above described active sensing implementation with dynamic camera configurations as opposed to just using fixed front-facing stereo cameras, i.e. the cameras are permanently in camera mode (a) of Fig. 1. Both methods were tested in two scenarios with different prior information. The experiment was repeated 10 times for each method and scenario.

The sequence of landmarks in the simulation environment consisted of  $N = 4$  landmarks. Different initial beliefs of the landmark positions with 50% confidence intervals can be seen in Fig. 3 with the actual landmark positions, and the initial position of the robot. The gaze direction of the cameras is  $40^\circ$  when looking left,  $-40^\circ$  when looking right, and  $0^\circ$  when facing forward. The angle of view of the cameras is  $\alpha = 50^\circ$  and maximum range  $r_{max} = 2.5$  m. The vision system outputs observations at 1 Hz rate. The noise parameters used were  $\sigma_v^2 = \sigma_\omega^2 = 0.01$  m<sup>2</sup>/s<sup>2</sup> and rad<sup>2</sup>/s<sup>2</sup> accordingly,  $\sigma_x^2 = 0.01$  m<sup>2</sup>,  $\sigma_y^2 = 0.02$  m<sup>2</sup>, and  $\sigma_\beta^2 = 0.05$  rad<sup>2</sup>. The number of samples to draw for the Monte Carlo sampling was  $K = 10$ . The weighing factors were  $\gamma = 0.7$ ,  $\xi = 0.3$ , and the cost  $\zeta = 0.01$ . The robot starts navigating to the next landmark when it believes to be closer than 0.2 m to the current landmark in sequence.

Fig. 3 presents typical outcomes of the simulations in both tested scenarios. In the case of Fig. 3a the robot with fixed cameras never reaches all of the landmarks because it fails to see the actual third landmark situated to the left of its originally planned trajectory. When the proposed



(a) With fixed front facing stereo cameras the robot fails to see the third landmark and continues towards the prior mean position. When operated with the active sensing enabled the robot observes the actual landmark and corrects its trajectory accordingly.

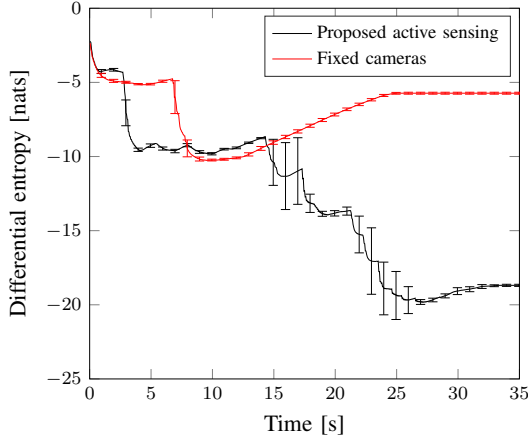


(b) When the prior means are close to the actual landmark positions, there is little difference in the performance of the two methods tested.

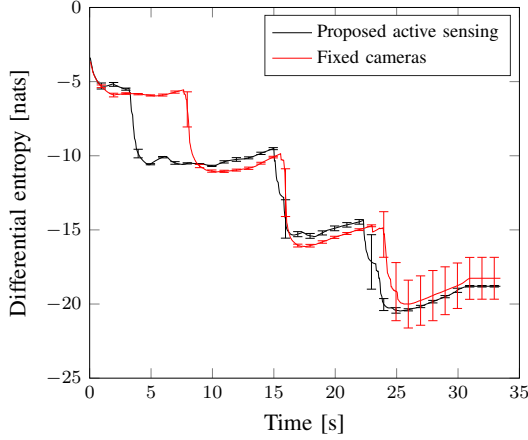
Fig. 3: Simulation experiments both with and without using the active sensing. The initial position of the robot is marked with the triangle. Circular markers denote the prior of the landmarks, with the associated dotted ellipses representing the 50% confidence intervals. Diamond markers denote the actual positions of the landmarks. The dashed line is the robot trajectory when active sensing is not used and the solid line is the robot trajectory when using the proposed active sensing.

active sensing is used instead, the robot keeps observing its surroundings in order to minimize the uncertainty about the landmark and upon observing the actual landmark, corrects its belief and plans a new trajectory to follow.

To quantify the results, the differential entropy of the



(a) Prior distribution as depicted in Fig. 3a.



(b) Prior distribution as depicted in Fig. 3b.

Fig. 4: Time evolution of belief state entropy. The lines indicate mean differential entropy over 10 experiments and the bars indicate 95% confidence intervals.

belief state during the experiments was studied and can be seen in Fig. 4a. Lower entropy indicates lower uncertainty. Using the proposed active sensing, the robot visited all the landmarks on each experiment, whereas the robot with fixed cameras never observed the third and fourth landmarks and only navigated to the prior mean locations. This can be seen also from the development of differential entropy; after the second sharp decline, where the robot first observes a new landmark, the differential entropy starts increasing due to the control noise, and finally reaches a steady state when the robot stops. With the proposed active sensing the differential entropy keeps declining when new landmarks are observed, and a much lower final value is reached.

If, however, the actual landmark positions are close to the prior means, there is little difference between the performance of the proposed active sensing and fixed cameras as seen in Fig. 3b. From the sharp declines of differential entropy of the belief state in Fig. 4b, we can see that the active sensing method intuitively observes the landmarks sooner than the method with fixed front-facing cameras. The

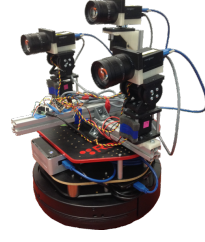


Fig. 5: Turtlebot equipped with the camera rig, which was used to carry out the experiments described in Section IV-B.

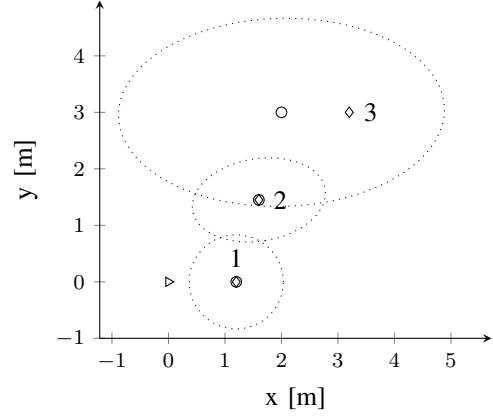


Fig. 6: Setting of the real world experiment. Initial position of the robot is marked with the triangle. Circular markers denote the prior of the landmarks, with the associated dotted ellipses representing the 50% confidence intervals. Diamond markers denote the actual positions of the landmarks.

fixed camera method, however, reaches lower differential entropy values after observing a landmark, because all three cameras are used as stereo pairs, whereas the active sensing methods usually only observe the landmark with one stereo pair, and operates the third camera in a monocular mode. There is no significant difference in the final entropy of the two methods. The fixed camera method failed to observe the fourth landmark in one of the experiments, which resulted in the large confidence intervals seen at the end.

Using the proposed active sensing is especially beneficial when the prior information about the landmark positions is highly uncertain and the actual landmark positions differ from the prior means. Random actions in the dynamic configuration could lead to similar results, but would be impractical and wear out a physical system.

#### B. Experiments on a real robot

The robot seen in Fig. 5 was used to demonstrate the developed algorithms on physical hardware. The robot features three Point Grey Grasshopper3 USB3 machine vision cameras with  $2048 \times 2048$  resolution mounted on a Turtlebot platform. The gaze direction of the cameras can be controlled between  $[-\frac{\pi}{2}, \frac{\pi}{2}]$  radians. The angle of view of the cameras is  $\alpha = 50^\circ$ . QR codes are used as landmarks since they are easy to recognize from the images and identify. Successfully

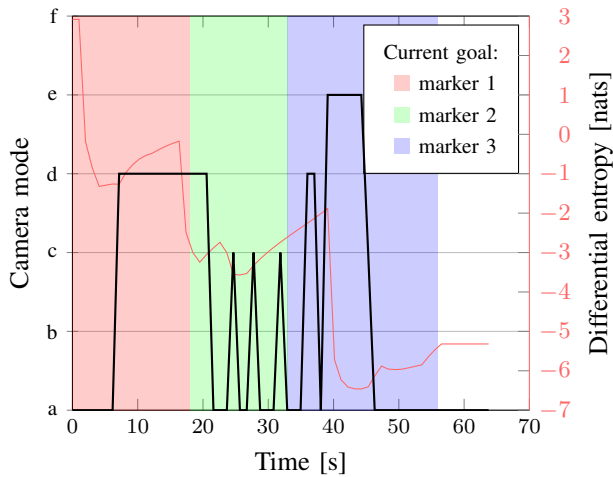


Fig. 7: Time evolution of used camera modes and the belief state entropy.

identifying the QR codes limits the maximum range to  $r_{max} = 4$  m. The on-board computer of the robot is an Intel NUC with a 3.1 GHz i7 processor and 8 GB of RAM.

The initial setting of the experiment is pictured in Fig. 6. The mean of the prior of the third landmark is off by so much, that the landmark is not observed by static front facing cameras. Behavior of the developed active sensing is visible in Fig. 7. At first, uncertainty of first landmark is reduced by directing all cameras towards it. After that, the algorithm prefers to keep one camera directed at the first landmark, and a stereo pair towards the second landmark. Between time-instances 21 and 39, the camera mode is frequently changed to observe the highly uncertain third landmark. After the observation is acquired, the cameras are directed at the landmark. Demonstration of the system in action can be found on <https://www.youtube.com/watch?v=n6AWOLpbNzs>.

## V. CONCLUSION

We studied an active sensing problem where three cameras are operated in six configurations, and used as either monocular cameras or stereo pairs. A robot follows a trajectory defined as a sequence of visual landmarks at uncertain locations and has to direct its gaze to maximize information about the landmark locations. The problem was formulated as a POMDP with a reward function based on mutual information and a myopic solution was provided.

Simulation experiments demonstrate the benefits of using the proposed active sensing implementation over using fixed front-facing stereo pairs. Our active sensing solution outperforms the fixed camera solution when prior information is highly uncertain and the prior mean differs from the actual landmark positions. If prior information is accurate or the prior means happen to be close to the actual landmark positions, there is little difference between the two tested methods. There was, however, no disadvantage in using the active sensing in any of the tests performed. Real world

experiments demonstrate that the algorithm runs real-time on physical hardware.

The experiments are not comprehensive and can be only consider proof-of-concept, but the method shows promising results that motivate further development. In future work, the solution will be generalized, and we plan to apply similar active sensing methods to a wider range of problems in mobile robotics.

## ACKNOWLEDGMENT

The authors would like to thank Mikko Lauri for his helpful comments and suggestions. The part by Risto Ritala was funded by the Academy of Finland, project “*Optimal operation of observation subsystems in autonomous mobile machines*”, which is gratefully acknowledged.

## REFERENCES

- [1] L. Johnson, B. Sullivan, M. Hayhoe, and D. Ballard, “Predicting human visuomotor behaviour in a driving task,” *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 369, no. 1636, 2014.
- [2] A. Hero and D. Cochran, “Sensor management: Past, present, and future,” *Sensors Journal, IEEE*, vol. 11, no. 12, pp. 3064–3075, Dec 2011.
- [3] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, “Planning and acting in partially observable stochastic domains,” *Artificial Intelligence*, vol. 101, no. 12, pp. 99 – 134, 1998.
- [4] C. Stachniss, G. Grisetti, and W. Burgard, “Information gain-based exploration using Rao-Blackwellized particle filters,” in *Robotics: Science and Systems*, 2005.
- [5] M. Spaan and P. Lima, “A decision-theoretic approach to dynamic sensor selection in camera networks,” in *19th International Conference on Automated Planning and Scheduling (ICAPS-2009)*, Sep 19 2009.
- [6] M. Spaan, T. Veiga, and P. Lima, “Decision-theoretic planning under uncertainty with information rewards for active cooperative perception,” *Autonomous Agents and Multi-Agent Systems*, pp. 1–29, 2014.
- [7] M. Araya-López, O. Buffet, V. Thomas, and F. Charpillet, “A POMDP extension with belief-dependent rewards,” in *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2010, vol. 23, pp. 64–72.
- [8] L. Itti and P. Baldi, “Bayesian surprise attracts human attention,” *Vision research*, vol. 49, no. 10, pp. 1295–1306, 2009.
- [9] V. Krishnamurthy and D. Djonin, “Structured threshold policies for dynamic sensor scheduling—a partially observed Markov decision process approach,” *Signal Processing, IEEE Transactions on*, vol. 55, no. 10, pp. 4938–4957, Oct 2007.
- [10] R. Martinez-Cantin, N. de Freitas, E. Brochu, J. Castellanos, and A. Doucet, “A Bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot,” *Autonomous Robots*, vol. 27, no. 2, pp. 93–103, 2009.
- [11] M. Lauri and R. Ritala, “Stochastic control for maximizing mutual information in active sensing,” in *ICRA 2014 Workshop: Robots in Homes and Industry: Where to Look First?*, 2014.
- [12] R. Bellman, *Dynamic Programming*. Princeton University Press, 1957.
- [13] M. L. Puterman, *Markov Decision Processes—Discrete Stochastic Dynamic Programming*. Hoboken, New Jersey: John Wiley & Sons, Inc., 1994.
- [14] D. White, *Markov Decision Processes*. John Wiley & Sons, Ltd., 1993.
- [15] R. D. Smallwood and E. J. Sondik, “The optimal control of partially observable Markov processes over a finite horizon,” *Operations Research*, vol. 21, no. 5, pp. 1071 – 1088, 1973.
- [16] S. Ross, J. Pineau, S. Paquet, and B. Chaib-draa, “Online planning algorithms for POMDPs,” *Journal of Artificial Intelligence Research*, vol. 32, pp. 663–704, 2008.
- [17] S. Kullback, *Information Theory and Statistics*. Dover Publications Inc., 1968.
- [18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006.